PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2003-069924
(43)Date of publication of application : 07.03.2003

(51)Int.Cl. H04N 5/76

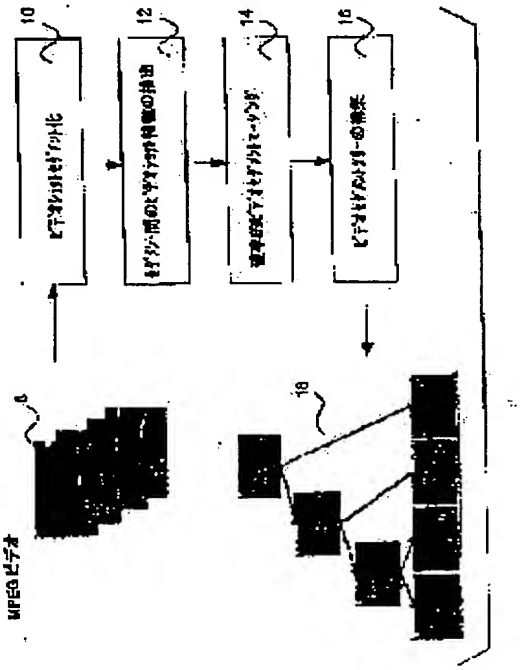
(21)Application number : 2002-230037
(22)Date of filing : 07.08.2002

(71)Applicant : EASTMAN KODAK CO
(72)Inventor : LOUI ALEXANDER C
GATICA-PEREZ DANIEL

(30)Priority
Priority number : 2001 927041 Priority date : 09.08.2001 Priority country : US

(54) VIDEO STRUCTURING METHOD BY PROBABILISTIC MERGING OF VIDEO SEGMENT

(57)Abstract:
PROBLEM TO BE SOLVED: To generate video clusters without the need for decision of empirical parameters.
SOLUTION: This invention provides the method for structuring video by probabilistic merging of video segments includes the steps of obtaining a plurality of frames of unstructured video; generating video segments from the unstructured video by detecting short boundaries based on color dissimilarity between consecutive frames; extracting a feature set by processing pairs of segments for visual dissimilarity and their temporal relationship, thereby generating an inter-segment visual dissimilarity feature and an inter-segment temporal relationship feature; and merging video segments with a merging criterion that applies a probabilistic analysis to the feature set, thereby generating a merging sequence representing the video structure. The probabilistic analysis follows a Bayesian formulation and the merging processing is represented in a hierarchical tree structure including frames extracted from each segment.



LEGAL STATUS

[Date of request for examination] 04.08.2005
[Date of sending the examiner's decision of rejection]
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date of final disposal for application]
[Patent number]
[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12)**公開特許公報**（Ａ）　(11)特許出願公開番号
特開2003－69924
（P2003－69924A）
(43)公開日　平成15年３月７日(2003.3.7)

(51)Int.Cl.'	識別記号	ＦＩ	ﾃｰﾏｺｰﾄﾞ(参考)
Ｈ０４Ｎ　5/76		Ｈ０４Ｎ　5/76	Ｂ　５Ｃ０５２

審査請求　未請求　請求項の数３　ＯＬ　（全 15 頁）

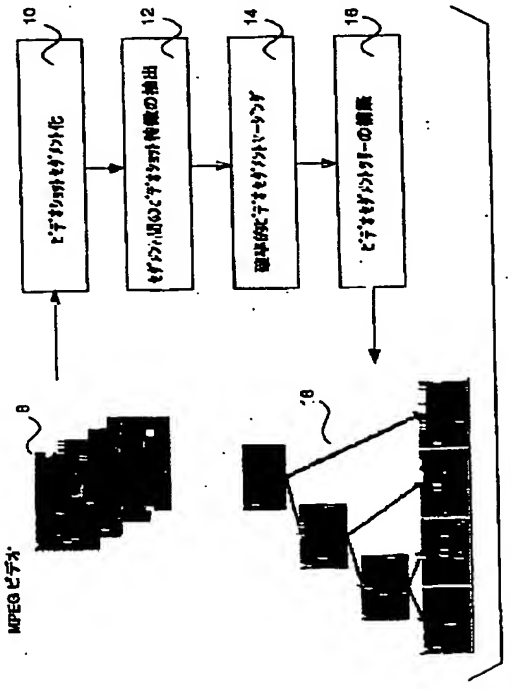
(21)出願番号　特願2002－230037(P2002－230037)
(22)出願日　平成14年８月７日(2002.8.7)
(31)優先権主張番号　９２７０４１
(32)優先日　平成13年８月９日(2001.8.9)
(33)優先権主張国　米国（ＵＳ）

(71)出願人　590000846
イーストマン　コダック　カンパニー
アメリカ合衆国，ニューヨーク14650，ロチェスター，ステイト　ストリート343
(72)発明者　アレクサンダー　シー　ルイ
アメリカ合衆国　ニューヨーク　14526
ペンフィールド　セラマー・ドライヴ　8
(74)代理人　100070150
弁理士　伊東　忠彦　（外３名）

最終頁に続く

(54)【発明の名称】　ビデオセグメントの確率的マージングによるビデオ構造化方法

(57)【要約】
【課題】経験的なパラメータ決定を要せずビデオクラスタの生成を可能とすること。
【解決手段】　ビデオセグメントの確率的マージングによりビデオを構造化する方法であって、構造化されていない複数のビデオのフレームを取得するステップと、上記構造化されていないビデオから、連続するフレーム間の色の非類似度に基づきショットの境界を検出することにより、ビデオセグメントを生成するステップと、セグメント間の可視的な非類似度特徴及びセグメント間の時間的関係特徴を生成するため、対のセグメントを処理することにより、可視的な非類似度及びそれらの時間的な関係に対する特徴セットを抽出するステップと、ビデオ構造を表現するマージングシーケンスを生成するため、上記特徴セットに確率的な解析を適用するマージング判断基準によりビデオセグメントをマージングするステップとを含む。確率的解析は、ベイズの定式化に後続し、マージングシーケンスは、各セグメントから抽出されたフレームを含む階層的ツリー構造で表現される。



【特許請求の範囲】
【請求項１】　ビデオセグメントの確率的マージングによりビデオを構造化する方法であって、
ａ）構造化されていない複数のビデオのフレームを取得するステップと、
ｂ）連続するフレーム間の色の非類似度に基づきショットの境界を検出することによって、上記構造化されていないビデオからビデオセグメントを生成するステップと、
ｃ）対のセグメントを可視的な非類似度及び時間的な関係に対して処理して、セグメント間の可視的な非類似度特徴及びセグメント間の時間的関係特徴を生成することにより、特徴セットを抽出するステップと、
ｄ）上記特徴セットに確率的な解析を適用するマージング判断基準を用いてビデオセグメントをマージングし、ビデオ構造を表現するマージングシーケンスを生成するステップとを含む、方法。
【請求項２】　上記ステップｂ）は、連続するフレームからカラーヒストグラムを生成するステップと、
上記カラーヒストグラムから、連続するフレーム間の色の非類似度を表現する差分信号を生成するステップと、複数のフレームに亘り求めた平均非類似度に基づき上記差分信号を閾値処理し、これにより、ショットの境界の存在を特定する信号を生成するステップとを含む、請求項１記載の方法。
【請求項３】　上記差分信号は、上記連続するフレームの一のフレームを中心としビデオキャプチャのフレームレートの分数に対応するフレーム数を持つ複数のフレーム、に亘り求めた平均非類似度に基づく、請求項２記載の方法。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】本発明は、一般的には、ビデオ素材の処理及び検索に係り、より詳細には、ホームビデオからの情報にアクセスし、組織化し、操作することに関する。
【０００２】
【従来の技術】ビデオコンテンツのすべてのソースの中で、構造化されていない消費者のビデオは、おそらく、大部分の人々が現に若しくは最終的に処理することに関心を持ちうるコンテンツを構成している。ホームビデオにアクセスしそれを操作することにより、個人的な記憶を組織化し編集することは、従来の静止画像の組織化の自然な技術的拡張である。しかしながら、かかる取り組みは、デジタルビデオの出現に誘起されるものの、これらの可視的なアーカイブのサイズにより、及び、ホームビデオ情報にアクセスし、組織化し、操作するための効果的なツールの不足により、制限されたままである。かかるツールの作成は、アルバムやビデオベビーブック中のビデオイベントの組織化や、ビデオデータやマルチメディアファミリーウエブページから抽出した静止画によるポストカードの編集等へのドアを開放する。実際には、多種多様なユーザの興味により対話型の解決策が提案されており、これは、意味のあるレベルで所望のタスクを特定するために最小限のユーザフィードバックを必要とし、また、冗長であるか若しくは高い信頼性で実行できるタスクに対して自動化されたアルゴリズムを付与する。
【０００３】コマーシャルビデオにおいて、多くの動画文書は、可視的なコンテンツに反映されるストーリ構造を有する。かかる場合、完全な動画文書は、ビデオクリップと称される。ビデオの生成の基本単位は、ショットであり、連続的な動作を捕捉する。ビデオショットの特定は、各ショットの開始及び終了を与える風景変化検出スキームによって実現される。風景は、通常的には、位置若しくは印象的な出来事を基にして統合される、少数の相互関係のあるショットからなる。特集フィルムは、典型的には、動画文書のコンテンツを理解するためにストーリラインを定義する、多数の風景からなる。
【０００４】コマーシャルビデオとは対照的に、規制されていないコンテンツ及びストーリラインの不在が、ホームビデオの主なる特性である。消費者コンテンツは、通常的には、それぞれが時間軸上でランダムに広がる１若しくは２，３のショットから構成される、関連し若しくは独立したイベントのセットからなる。かかる特性により、消費者ビデオは、ストーリラインモデルに基づくビデオ解析アプローチに対して不適となる。しかし、可視的な類似性、及び、大型のホームビデオデータベースの統計的解析後に明確に表れるビデオセグメント間の時間的近接度（temporal adjacency）に基づく時空間（spatio-temporal）構造が依然として存在する。かかる構造は、消費者の静止画像の構造と実質的に等価であり、ホームビデオの構造化をクラスタリングの問題として対処することを示唆する。当面のタスクは、所与のビデオクリップに存在するクラスタ数の決定、及び、ビデオシーケンス内の各フレーム／ショットにクラスタラベルを割り当てるための最適な基準の設計として定義されうる。これは、実際に、ストーリラインされたコンテンツを処理するときでさえ、ビデオ解析の大部分の研究で取られる方向性である。
【０００５】例えば、米国特許第５８２１９４５号には、検索のための複雑なビデオ選択の階層的な分解を抽出し、ビデオの風景間及び風景内の重要な関係を捕捉するために可視的情報及び時間的情報を結合する技術が開示されている。従って、これは、コンテンツの先見的な知識を用いない潜在的なストーリ構造の解析を可能とする。かかるアプローチは、ビデオショット境界決定（ショットセグメント化）及びショットクラスタリングという、２段階の原理体系のバリエーションでビデオ構造化を実現する。第１段階は、これまでビデオ解析で最も研

究されている（例えば、U.Gargi, R.Kasturi and S.H.Strayer, "Performance Characterizationof Video-Shot-Change Detection Method", IEEE CSVT, Vol.10,No.1,February 2000,p1-13参照）。第２段階に対して、ビデオ構造の基本単位としてショットを用いて、ｋ－ｍｅａｎｓ法、分散をベースとしたクラスタリング、及び、時間制約付き（time-constrained）マージング技術のすべてが、先行技術に開示されている。これらの方法の幾つかは、通常的には、アプリケーションに依存するか若しくはユーザフィードバックにより経験的に決定される、多くのパラメータの設定を必要とする。

【０００６】

【発明が解決しようとする課題】先行技術で理解できるように、階層的な表現は、構造化されていないコンテンツを表現するのに自然であるばかりか、検索及び操作のための有用な非線形対話モデルを付与する最良な方法であるだろう。幸いなことに、副産物として、クラスタリングにより、ビデオコンテンツに対する階層的な表現の生成が可能となる。風景遷移グラフ（例えば、米国特許第５８２１９４５号参照）及びツリーに基づくコンテンツのテーブルを含む、階層的な組織化のための異なるモデルは、先行技術において提案されているが、各特定のモデルの効率／有用性は、一般的に、開かれた問題のままである。

【０００７】現在に至るまで、幾つかの研究しかホームビデオの解析を取り扱っていない（例えば、G.Iyengar and A.Lippman, "Content-based Browsing and Editio n ofUnstructured Video", IEEE ICME, New York City, August 2000; R.Lienhart, "Abstracting Home Video Automatically", ACM Multimedia Conference, Orlando, October, 1999,pp.37-41; Y.Rui and T.S.Huang, "A Unified Framework for Video Browsing and Retrieval", in A.C.Ed., Handbook of Image andVideo Processing, Academic Press,1999参照）。Lienhartの研究は、ビデオサマリーの生成のためのクラスタリングを実行するために時刻記録された（time-stamped）情報を用いる。しかしながら、時刻記録された情報は、常に利用可能ではない。デジタルカメラはこの情報を含んでいるものの、ユーザは時間オプションを常に使用するとは限らない。それ故に、一般的な解決策は、この情報に依存することができない。Rui及びHuangによる、非常に簡易な統計的仮定に基づくコンテンツのテーブルの生成のための研究は、"ストーリライン"を備えた幾つかのホームビデオで試験を行った。しかしながら、ホームビデオの持つ非常に高いレベルの構造化されていない特性により、特別なストーリラインモデルのアプリケーションが極めて制限される。Iyengar及びLippmanによる記述を除いて、上述のアプローチのいずれも、かかるコンテンツの特有の統計データを詳細に解析していない。この観点から、本発明は、ショット持続時間の統計的モデルに基づくショット境界検出のためのベイズの法則を提案するN.Vasconcelos及びA.Lippmanによる"A Bayesian Video Modeling Framework for Shot Segmentation and Content Characterization",Proc.CVPR,1997の研究、及び、異なる確率法則を用いてホームビデオ解析を取り扱うIyengar及びLippmanによる研究により関連している。

【０００８】それにも拘らず、先行技術からは、ショットを組織化の基本単位として使用する確率的な原理体系が、対話のためのビデオ階層の生成を支援できるのかが不明である。本発明に至る際、消費者ビデオにおける可視的及び時間的な特徴の統計モデルが組織化のために詳細に調べられた。特に、ベイズの法則は、ホームビデオの時空間構造の事前知識をコード化するのに適するように思われた。先行技術を出発点として、ここで言及する画期的アプローチは、可視的な類似性、時間的な近接度及び持続時間のセグメント間の特性をジョイントモデルで統合し、経験的なパラメータ決定を用いずにビデオクラスタの生成を可能とする、効果的な確率的ビデオセグメント・マージング・アルゴリズムに基づく。

【０００９】本発明は、上述の問題点の一若しくはそれ以上を克服することを目的とする。

【００１０】

【課題を解決するための手段】簡潔に概説すると、本発明による一局面によれば、ビデオセグメントの確率的マージングによりビデオを構造化する方法であって、

ａ）構造化されていない複数のビデオのフレームを取得するステップと、

ｂ）上記構造化されていないビデオから、連続するフレーム間の色の非類似度に基づきショットの境界を検出することによりビデオセグメントを生成するステップと、

ｃ）セグメント間の可視的な非類似度特徴及びセグメント間の時間的関係特徴を生成するため、対のセグメントを処理することにより、それらの可視的な非類似度及び時間的な関係に対する特徴セットを抽出するステップと、

ｄ）ビデオ構造を表現するマージングシーケンスを生成するため、上記特徴セットに確率的な解析を適用するマージング判断基準によりビデオセグメントをマージングするステップとを含む、方法が提供される。好ましい実施例では、確率的解析は、ベイズの定式化に後続し、マージングシーケンスは、各セグメントから抽出されたフレームを含む階層的ツリー構造で表現される。

【００１１】上述の如く、本発明は、確率的モデルに基づいて消費者ビデオを構造化する方法を採用する。より具体的には、本発明は、ビデオショットを組織化の基本単位として用いて、ホームビデオ内のクラスタ構造を発見する新規な原理体系を提案する。本原理体系は、２つの概念に基づく。即ち、（ｉ）セグメント間の可視的な類似度、及び、時間的近接度とホームビデオセグメントの持続時間とを含むセグメント間の時間的な関係を表わ

すための統計モデル（例えば、学習済みジョイント混合ガウスモデル）の改良、及び、（ｉｉ）連続的なバイナリクラス化処理として階層的クラスタリング（マージング）の再定式化である。モデルは、確率的クラスタリングアルゴリズムにて上記（ｉｉ）で使用され、これに対して、ベイズの定式化は、これらのモデルがホームビデオの統計的な構造の事前知識を組み込むことができるので有用であり、原理化された方法体系の効果を供する。

【００１２】ビデオ構造化アルゴリズムは、本発明により効率的に実行可能であり、如何なる特別なパラメータ決定を必要としない。付随的に、ビデオクラスタを見出すことは、ビデオコンテンツに対する階層的な表現の生成を可能とし、検索及び操作のための非線形的なアクセスを提供する。

【００１３】本発明の主なる効果は、クラスタ検出及び個々のショットクラスタラベリングに関する本原理体系の性能に基づいて、消費者ホームビデオにおいて見出されるだろう構造化されていないコンテンツを有するビデオ及び構造化されていないビデオを処理することができることである。従って、それは、対話的な組織化及びホームビデオ情報の抽出のためのシステム用ツールを確立するための第１ステップである。

【００１４】ベイズのビデオセグメントマージングアルゴリズムに基づく消費者ビデオ構造化のための原理体系として、その他の効果は、本方法が、経験的なパラメータ決定を用いることなく、マージング処理を自動的に支配し、また、可視的及び時間的なセグメント非類似度特徴を単一モデルに統合することである。

【００１５】更に、ツリーによるマージングシーケンスの表現は、ビデオコンテンツへの階層的な非線形のアクセスを可能とするユーザインターフェースのための基礎を提供する。

【００１６】本発明のこれらの及び他の局面、目的、特徴及び効果は、好ましい実施例の次の詳細な説明及び上記特許請求の範囲を精査し、添付図面を参照することにより、より明確に理解できるだろう。

【００１７】

【発明の実施の形態】ショット検出及びクラスタ解析を用いるビデオ処理システムは公知であるので、本説明は、本発明によるビデオ構造化技術を形成し若しくはより直接的に協働する構成に特に向けられる。ここで具体的に開示されない構成は、本分野で知られている構成から選択されてよい。次の説明では、本発明の好ましい実施例が、通常的にはソフトウェアプログラムとして実行されるだろうが、当業者であればかかるソフトウェアの均等物がハードウェア内に構築されてよいことを容易に理解するだろう。次に開示される本発明によるシステムにおいて、本発明の実現のために有用であるがここで具体的に開示・提案・言及されない素材、ソフトウェアは、従来的であり当業者の通常的な設計事項である。本発明がコンピュータープログラムとして実現される場合、プログラムは、例えば磁気ディスク（フロッピー（登録商標）ディスクやハードドライブのような）や磁気テープのような磁気記憶媒体、光ディスク、光テープや機械読取り可能なバーコードのような光記憶媒体、ランダムアクセスメモリ（ＲＡＭ）やリードオンリーメモリ（ＲＯＭ）のような固体電子記憶デバイス、コンピュータープログラムを記憶するために用いられる他の物理的デバイスやメディアのような、従来的なコンピュータ一読取り可能な記憶媒体に記憶されてよい。

【００１８】ホームビデオに記憶された個人的な記憶に対してアクセス、組織化及び操作を行うことは、その構造化されていないコンテンツ及び明確なストーリライン構造の不足に起因して、技術的な課題となる。本発明において、原理体系は、消費者ビデオクリップ中の可視的情報の単位であるショット間の類似度及び近接度の統計的なパラメトリックモデルの発展型に基づく。ショットをマージング（統合）するベイズの法則は、これらのモデルがホームビデオの統計的構造の事前知識をコード化できるので、理にかなった選択に思われる。それ故に、本原理体系は、ショット境界検出及びベイズのセグメントマージングに基づく。セグメント間の可視的な類似度、時間的な近接度、及び、ＥＭアルゴリズムを用いてホームビデオ訓練サンプルから学習されたセグメント持続時間、のガウス混合ジョイントモデルは、観察した特性のクラス条件付き密度を表わすために用いられる。かかるモデルは、バイナリベイズクラス分類器（binary Bayes classifier）を構成するマージングアルゴリズムで使用され、このとき、マージングオーダーは、Highest Confidence First（ＨＣＦ）の変形により決定され、Maximum a Posteriori（ＭＡＰ）判定基準がマージング判定基準を定める。マージングアルゴリズムは、階層的なキューの使用により効率的に実行され、如何なる経験的なパラメータ決定を必要としない。最終的に、ツリーによるマージングシーケンスの表現がユーザインターフェースのための基礎を付与し、ビデオコンテンツへの階層的で非線形のアクセスを可能とする。

【００１９】先ず図１を参照するに、ビデオ構造化方法は、消費者ホームビデオで見出されるような、制約のないコンテンツを典型的に表示する構造化されていないビデオソースから得られる一連のビデオフレーム段階８に対して動作するように示されている。本発明によるビデオ構造化方法の主な特徴は、次の４つの段階で簡潔に要約される（後の段落で詳説する）。

【００２０】１）ビデオセグメント化段階１０：ショット検出がヒストグラム差分信号の適応的閾値処理（adaptive thresholding）により計算される。１－Ｄカラーヒストグラムは、各バンドに対してＮ＝６４量子化レベルにより、ＲＧＢ空間で計算される。Ｌ１メトリックが、２つの連続するフレーム間の非類似度ｄｃ（ｔ、ｔ

＋１）を表現するために使用される。後処理ステップとして、インプレース・モフォロジック（形態学的な）ｈｉｔ－ｏｒ－ｍｉｓｓ変換（in-place morphological hit-or-miss transform）が、複数の近接ショット境界の存在を無くす対の構造化要素によりバイナリ信号に施される。

【００２１】２）ビデオショット特徴抽出段階１２：可視的な類似度が２つの異なるビデオイベントの区別するのに十分でないことは本分野で公知である（Rui及びHuangの記述を参照）。可視的な類似度及び時間的な情報の双方は、先行技術においてショットクラスタリングのために使用されている（しかしながら、かかる変形の統計学的な特性は、ベイズの観点の下で研究されていない）。本発明では、ビデオシーケンスの３つの主なる特徴が、後のマージングに対する判断基準として利用される：即ち、

・可視的な類似度は、セグメント外観を表わす平均セグメントヒストグラムにより表される。平均ヒストグラムは、セグメント内の支配的な色の存在及びそれらの持続性（パーシステンス）の双方を表現する。

・セグメント間の時間的な分離は、同一のクラスタへのそれらの帰属の強力な指標である。

・２つの個々のセグメントの合計の持続時間も、同一のクラスタへのそれらの帰属に関する強力な指標である（例えば、２つの長いショットは、同一のビデオクラスタに属する可能性が小さい）。

【００２２】３）ビデオセグメントマージング段階１４：このステップは、ベイズ決定理論に基づく２クラス（マージ／マージなし）パターンクラス分類器を構成することにより実行される。セグメント間の可視的な類似度、時間的な近接度及びＥＭアルゴリズムを用いてホームビデオ訓練サンプルから学習されたセグメント持続時間のガウス混合ジョイントモデルは、観察した特性のクラス条件付き密度を表わすために用いられる。かかるモデルは、バイナリベイズクラス分類器（binary Bayes classifier）を構成するマージングアルゴリズムで使用され、このとき、マージングオーダは、Highest Confidence First（ＨＣＦ）の変形により決定され、Maximum a Posteriori（ＭＡＰ）判定基準がマージング判定基準を定める。マージングアルゴリズムは、階層的なキューの使用により効率的に実行され、如何なる経験的なパラメータ決定を必要としない。マージング手順のフローチャートは、図２に示され、後に詳説される。

【００２３】４）ビデオセグメントツリー構築段階１６：マージングシーケンス、即ち、対のビデオセグメントの連続的なマージングを備えたリストは、階層を生成するために記憶及び使用され、そのマージングシーケンスは、バイナリ分割ツリー１８により表現される。図５は、典型的なホームビデオからのツリー表現を示す。

【００２４】１．本アプローチの概説

ビデオセグメントに対する特徴ベクトル表現が想定される。即ち、ビデオクリップがショット若しくはセグメント（セグメントは、１以上のショットからなる）に分割されており、それらを表現する特徴が抽出されていることを想定する。如何なるクラスタリング手順であっても、ホームビデオクリップ内の各セグメントにクラスタラベルを割り当てると共にクラスタ数（クラスタは一若しくはそれ以上のセグメントを包囲してよい）を決定するためのメカニズムを規定すべきである。クラスタリング処理は、ビデオイベントが限られた持続時間であるので（Rui及びHuangの記述を参照）、時間を制約として含む必要がある。しかしながら、ホームビデオのセグメント間の特徴に対する一般的な発生モデルの定義は、それらの構造化されていないコンテンツに起因して、特に困難である。代わりに、本発明によれば、ホームビデオは、統計的セグメント間モデルを用いて解析される。換言すると、本発明は、対のセグメントに定義された可視的及び時間的特徴の特性を表わすモデルを構築することを提案する。セグメント間の特徴は、必然的にマージングフレームワークに出現し、可視的な非類似度、持続時間及び時間的な近接度を統合する。マージングアルゴリズムは、対のビデオセグメントを連続的に取得し、それらがマージされるべきか否かを決定する、クラス分類器とみなすことができる。s_i及びs_jを、ビデオクリップ内のｉ番目のビデオセグメント及びｊ番目のビデオセグメントとし、εを、かかる対のセグメントが同一のクラスタに対応するか否か及びマージされるべきか否かを指示する、バイナリランダム変数（ｒ．ｖ．）とする。マージング処理の構成は、連続的な２クラス（マージ／マージなし）パターンクラス化問題として、ベイズ決定理論からの概念の適用を可能とする（ベイズ決定理論の詳細については、例えばR.O.Duda, P.E.Hart and D.G.Stork, Pattern Classification, 2th ed., John Wiley and Sons,2000を参照）。Maximum a Posteriori（ＭＡＰ）判定基準は、ｒ．ｖ．ｘ（セグメント間の特徴を表わし、後に詳説される。）のｎ次元の実現x_{ij}を仮定して、選択されなければならないクラスが、ｘの下でのεの事後確率質量関数を最大にするクラスであることを確立する。即ち、

【００２５】

【数１】

$$\varepsilon^* = \arg\max_{\varepsilon} \Pr(\varepsilon \mid x)$$

ベイズの法則により、

【００２６】

【数２】

$$\Pr(\varepsilon \mid x) = \frac{p(x \mid \varepsilon)\Pr(\varepsilon)}{p(x)}$$

ここで、ｐ（ｘ｜ε）はεの下でのｘの尤度であり、Ｐ

ｒ（ε）は、εの事前確率（prior）であり、Ｐ（ｘ）は、特徴の分散である。ＭＡＰ原理の適用は、次のように表現でき、

【００２７】
【数３】

$$\varepsilon^* = \begin{cases} 1, & p(x \mid \varepsilon = 1)\Pr(\varepsilon = 1) > p(x \mid \varepsilon = 0)\Pr(\varepsilon = 0) \\ 0 & otherwise \end{cases}$$

若しくは、標準的な仮説試験表記法では、ＭＡＰ原理は、次のように表現できる。即ち、

【００２８】
【数４】

$$p(x \mid \varepsilon = 1)\Pr(\varepsilon = 1) \underset{H_0}{\overset{H_1}{\gtrless}} p(x \mid \varepsilon = 0)\Pr(\varepsilon = 0)$$

ここで、H_1は、対のセグメントがマージされるべきであるとする仮説を指示し、H_0は、その逆を指示する。この公式を用いると、対のショットのクラス化は、所定の終了条件が満たされるまで連続的に実行される。それ故に、タスクは、有用な特徴空間の決定、分散に対するモデルの選択及びマージングアルゴリズムの規定である。これらのステップのそれぞれは、次に段落に言及される。

【００２９】２．ビデオセグメント化
基本のセグメントを生成するため、ショット境界検出が、段階１０で、ホームビデオに通常的に見出される切れ目（カット）を検出するための一連の方法により計算される（例えば、U.Gargi, R.Kasturi and S.H.Strayer, “PerformanceCharacterization of Video-Shot-Change Detection Method”, IEEE CSVT, Vol.10,No.1,February 2000,p1-13参照）。検出エラーに起因した過剰なセグメント化は、クラスタリングアルゴリズムにより対処できる。更に、中身のほとんどないビデオは除去される。

【００３０】本発明の好ましい実施例において、ショット検出は、ヒストグラム差分信号の適応的閾値処理により決定される。１－Ｄ（一次元）カラーヒストグラムは、各バンドに対してＮ＝６４量子化レベルにより、ＲＧＢ空間で計算される。他のカラーモデル（ＬＡＢやＬＵＶ）も使用可能であり、より良好なショット検出を付与しうるが、計算負荷が増加する。Ｌ１メトリックが、２つの連続するフレーム間の非類似度ｄｃ（ｔ、ｔ＋１）を表現するために使用される。即ち、

【００３１】
【数５】

$$d_C(t,t+1) = \sum_{k=1}^{3N} \left| h_t^k - h_{t+1}^k \right|$$

ここで、$h_t{}^k$は、フレームｔの連結ＲＧＢヒストグラムに対するｋ番目のビンの値を示す。次いで、一次元信号d_cは、ｆｒをフレームレートとしたとき、長さｆｒ／２の時間ｔで中心化したスライディング・ウインドウ（sliding window）により計算される閾値により二値化される。即ち、

【００３２】
【数６】

$$s(t) = \begin{cases} 1 & d_C(t) > \mu_d(t) + k\sigma_d(t) \\ 0 & otherwise \end{cases}$$

ここで、μ_d（ｔ）は、スライディング・ウインドウにより計算される非類似度の平均であり、σ_d（ｔ）は、その平均値周辺のデータセットの変動のよりロバストな推定量であることが知られている、ウインドウ内の非類似度の平均絶対偏差を示し、ｋは、閾値の決定のための信頼区間を設定する係数であり、閾値は、当該区間内に設定される。連続したフレームは、それ故に、ｓ（ｔ）＝０の場合には同一のショットに属すると考えられ、隣接するフレーム間のショット境界はｓ（ｔ）＝１のときに特定される。

【００３３】後処理ステップとして、インプレース・モフォロジック（形態学的な）ｈｉｔ－ｏｒ－ｍｉｓｓ変換（in-place morphological hit-or-miss transform）が、複数の近接ショット境界の存在を無くす対の構造化要素によりバイナリ信号に適用される。即ち、

【００３４】
【数７】

$$b(t) = s(t) \otimes (e_1(t), e_2(t))$$

ここで、

【００３５】
【外１】

$$\otimes$$

は、ｈｉｔ－ｏｒ－ｍｉｓｓを示し、構造化要素のサイズは、ホームビデオショット持続時間ヒストグラム（ホームビデオショットは、２秒より短く持続しないだろう）に基づき、ｆｒ／２に設定される（Jean Serra: Image Analysis and Mathematical Morphology, Vol.1, Academic Press, 1982参照）。

【００３６】３．ビデオセグメント間の特徴決定
可視的な非類似度、時間的分離度及び積算したセグメント持続時間に対する特徴セットは、ビデオショット特徴抽出段階１２で生成される。可視的な非類似度及び時間的情報の双方、特に時間的分離度は、過去にクラスタリングのために使用されている。可視的な非類似度の場合、可視的特徴の違いの分かる出力の観点で、単一のフ

レームが、セグメントのコンテンツを表現するのに十分でない場合があることは明らかである。幾つかの利用可能な解決法から、平均セグメントカラーヒストグラムが、セグメント外観を表現するために選択される。即ち、

【００３７】

【数８】

$$m_i = \frac{1}{M_i}\sum_{t=b_i}^{e_i} h_t$$

ここで、h_tは、ｔ番目のカラーヒストグラムを示し、m_iは、それぞれ$M_i = e_i - b_i + 1$のフレーム（e_i及びb_iは、セグメントs_iの開始及び終了フレームを示す）からなるセグメントs_iの平均ヒストグラムを示す。平均ヒストグラムは、セグメント内の支配的な色の存在及びそれらの持続性を表現する。平均セグメントヒストグラム差のＬ１ノルムが対のセグメントｉ，ｊを実質的に比較するために用いられる。即ち、

【００３８】

【数９】

$$\alpha_{ij} = \sum_{k=1}^{B} | m_{ik} - m_{jk} |$$

ここで、α_{ij}は、セグメントｉ，ｊ間の可視的非類似度を示し、Ｂはヒストグラムビンの数を示し、m_{ik}は、セグメントs_iの平均カラーヒストグラムのｋ番目のビンの値であり、m_{jk}は、セグメントs_jの平均カラーヒストグラムのｋ番目のビンの値である。

【００３９】時間的情報の場合、同一のクラスタへのそれらの帰属についての強力な指標である、s_iとs_jとの間の時間的分離度は、次の通り定められる。即ち、

【００４０】

【数１０】

$$\beta_{ij} = \min(|e_i - b_j|, |e_j - b_i|)(1 - \delta_{ij})$$

ここで、δ_{ij}は、クロネッカーのデルタを示し、b_i，e_iは、セグメントs_iの最初と最後のフレームを示し、b_j，e_jは、セグメントs_jの最初と最後のフレームを示す。

【００４１】更に、２つの別個のセグメントの積算セグメント（合計の）継続時間は、同一のクラスタへのそれらの帰属についての強力な指標である。図３は、グラウンドトルースによるデータベースからの約６６０個のショットに対する経験的なホームビデオショット持続時間を示し、その適合はガウス混合モデル（次の段落を参照）による。（図３には、経験的な持続時間、及び、６つの構成要素からなる推定されたガウス混合モデルが重ね合わせられている。持続時間は、最も長い持続時間（５８０秒）で正規化されている。）ビデオが異なるシナリオに対応し、複数の人によって撮影されているにも拘らず、明らかな時間的パターンが存在する（Vasconcelos及びLippmanの記述参照）。積算したセグメント持続時間τ_{ij}は、次の通り定義される。即ち、

【００４２】

【数１１】

$$\tau_{ij} = card(s_i) + card(s_j)$$

ここで、ｃａｒｄ（ｓ）は、セグメントｓ中のフレーム数を示す。

【００４３】４．尤度及び事前確率のモデリング

セグメント間の特徴セットの統計的モデリングは、ビデオセグメントマージング段階１４で生成される。上述の３つの特徴は、ベクトルｘ＝（α、β、τ）により、特徴空間Ｘの成分となる。２つのクラスの分離度を解析するため、図４は、ホームビデオから抽出された４０００個のラベル標示されたセグメント間特徴ベクトルの散在するプロットを示す。（サンプルの半分は、仮説H_1に対応し（セグメントペアが共に帰属し、薄い灰色でラベル標示）、他の半分は、仮説H_0に対応する（セグメントペアは、共に帰属せず、濃い灰色でラベル標示）。特徴は正規化されている。）

プロットは、２つのクラスが一般的に分離されていることを示す。このプロットの投影図は、純粋な可視的類似度に依存することの限界を明示する。パラメトリック混合モデルは、観察したセグメント間の各特徴クラス条件付き密度に対して適合される。即ち、

【００４４】

【数１２】

$$p(x|\varepsilon,\Theta) = \sum_{i=1}^{K_\varepsilon} \Pr(c=i)\, p(x|\varepsilon,\theta_i)$$

ここで、K_εは、各混合中の成分数を示し、Ｐｒ（ｃ＝ｉ）は、ｉ番目の成分の事前確率を示し、ｐ（ｘ｜ε、θ_i）は、θ_iによりパラメータ表示されるｉ番目のｐｄｆであり、すべてのパラメータのセットを表わす。本発明において、ｄ次元の混合の成分に対して下式の多変量ガウス形が想定され、

【００４５】

【数１３】

$$p(x|\varepsilon,\theta_i) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} e^{-\frac{1}{2}(x-\mu_i)^t \Sigma_i^{-1}(x-\mu_i)}$$

パラメータθ_iは、平均μ_i及び共分散行列（covariance matrices）Σ_iである（Duda他、Pattern Classification, op .cit.参照）。

【００４６】公知のＥＭアルゴリズムは、パラメータΘのセットの最尤（ＭＬ）推定法に対する標準の手順を構成する（A.P.Dempster, N.M.Laird and D.B.Rubin, “Maximum Likelihood from Incomplete Data via the EM Algorithm”, Journal of the Royal Statistical Society, Series B, 39:1-38,1977参照）。ＥＭは、観察データがある意味で不完全である場合に、問題の広い範囲に対するＭＬ推定値を見出すための知られた技術である。ガウス混合の場合、不完全データは、観察されない

混合成分であり、その事前確率はパラメータ｛Ｐｒ（ｃ）｝である。ＥＭは、反復的なヒルクライミング（hill-climbing）手順を用いることによる、観察データの下での完全データの対数尤度の条件付き期待値を増大することに基づく。更に、モデル選択、即ち各混合の成分数は、最大記述長（ＭＤＬ）原理を用いて自動的に推定できる（J.Rissanen, "Modeling by Shortest Data Description", Automatica, 14:465-471,1978）。

【００４７】一般的なＥＭアルゴリズムは、任意の分散に対して有効であり、観察データＸ＝｛x_1，…，x_n｝の下での完全データＹの対数尤度の条件付き期待値を増大することに基づく。即ち、

【００４８】

【数１４】

$$Q(\theta \mid \theta^{(p)}) = E\{\log p(Y \mid \theta) \mid x, \theta^{(p)}\}$$

であり、反復的なヒルクライミング手順を用いることによる。前式において、Ｘ＝ｈ（Ｙ）は、既知の多対一関数（many-to-one function）（例えば、部分集合演算子）を表わし、ｘはデータのシーケンス若しくはベクトルを表わし、ｐは反復数を示す上付き文字である。ＥＭアルゴリズムは、最大のＱ（θ）に収束するまで、次の２つのステップを繰り返す。即ち、

E-STEP：完全データの所期の尤度をθ、Ｑ（θ｜θ（ｐ））の関数として導出する。

M-STEP：次式により、パラメータを再推定する。

【００４９】

【数１５】

$$\theta^{(p+1)} = \arg\max_{\theta} Q(\theta \mid \theta^{(p)})$$

換言すると、第１にE-STEPで不完全データを埋める値を推定する（対数尤度自体に代わって、観察データの下での完全データの対数尤度の条件付き期待値を用いる）。次いで、M-STEPで用いる最大尤度パラメータ推定値を計算し、適切な終了基準に達するまで反復される。ＥＭは、サンプルセットの尤度の局所最大値に収束する反復アルゴリズムである。

【００５０】多変量ガウスモデルの特別の場合に対して、完全データはＹ＝（Ｘ，Ｉ）により与えられ、ここで、Ｉは観察データの各サンプルを生成する際に使用されたガウス成分を示す。要素に関しては、ｙ＝（ｘ，ｉ），ｉ＝｛1_1，…，K_ε｝である。かかる場合、ＥＭは、更に簡略化された形となる。即ち、

E-STEP：すべてのＮ個の訓練用サンプル及びすべての混合成分に対して、ガウスｉが現在の推定の下でサンプルx_jに適合する確率を計算する。

【００５１】

【数１６】

$$p(i \mid x_j, \varepsilon, \Theta^{(p)}) = \frac{\pi_i p(x_j \mid \varepsilon, \theta_i^{(p)})}{\sum_{k=1}^{K_\varepsilon} \pi_k p(x_j \mid \varepsilon, \theta_k^{(p)})}$$

M-STEP：パラメータを再推定する。

【００５２】

【数１７】

$$\pi_i^{(p+1)} = \frac{1}{N} \sum_{j=1}^{N} p(i \mid x_j, \varepsilon, \Theta^{(p)})$$

$$\mu_i^{(p+1)} = \frac{\sum_{j=1}^{N} x_j \, p(i \mid x_j, \varepsilon, \Theta^{(p)})}{\sum_{j=1}^{N} p(i \mid x_j, \varepsilon, \Theta^{(p)})}$$

$$\Sigma_i^{(p+1)} = \frac{\sum_{j=1}^{N} p(i \mid x_j, \varepsilon, \Theta^{(p)})(x_j - \mu_i^{(p+1)})(x_j - \mu_i^{(p+1)})^T}{\sum_{j=1}^{N} p(i \mid x_j, \varepsilon, \Theta^{(p)})}$$

混合成分のそれぞれに対する平均ベクトル及び共分散行列は、最初に初期化されなければならない。本実施例では、平均値は、従来的なＫ－ｍｅａｎｓアルゴリズムを用いて初期化されるが、共分散行列は、恒等行列を用いて初期化される。他のヒルクライミング法として、データ駆動型初期化は、通常的には、純粋なランダム初期化よりも良好に実現する。更に、ＥＭ反復の連続的な再始動時、少量のノイズが各平均値に加えられ、局所最大にトラップさせるための手順を漸減する。

【００５３】収束判定基準は、連続的な反復での観察データの対数尤度の増加率により規定される。

【００５４】

【数１８】

$$\log L(\Theta \mid X) = \log \prod_{j=1}^{N} p(x_j \mid \varepsilon, \Theta)$$

即ち、ＥＭ反復は、

【００５５】

【数１９】

$$\frac{\log L(\Theta^{(p+1)} \mid X) - \log L(\Theta^{(p)} \mid X)}{\log L(\Theta^{(p)} \mid X)} \leq 10^{-2}$$

のときに終了する。

【００５６】特別なモデル、即ち各混合の成分の数Ｋ$_{\varepsilon}$は、最大記述長（ＭＤＬ）原理を用いて、

【００５７】

【数２０】

$$K_{\varepsilon}^{*} = \arg\max_{K_{\varepsilon}} \left(\log L(\Theta \mid X) - \frac{n_{K_{\varepsilon}}}{2} \log N \right)$$

を選択することにより、自動的に推定される。ここで、Ｌ（・）は訓練用セットの尤度を示し、ｎ$_{Ｋ\ \varepsilon}$は、モデルに必要なパラメータ数であり、ガウス混合に対しては、

【００５８】

【数２１】

$$n_{K_{\varepsilon}} = (K_{\varepsilon} - 1) + K_{\varepsilon} d + K_{\varepsilon} \frac{d(d+1)}{2}$$

となる。２つのモデルが同様な方法でサンプルに適合するとき、より簡略化したモデル（より小さいＫ$_{\varepsilon}$）が選択される。

【００５９】変数の間で独立想定を課すことに代わって、完全なジョイントクラス条件付ｐｄｆｓが推定される。ｐ（ｘ｜ε＝０）及びｐ（ｘ｜ε＝１）に対するパラメトリックモデルのＭＬ推定は、上述の手順によって、それぞれの場合で１０個の成分により表現される確率密度を生成する。

【００６０】ベイズのアプローチでは、事前確率質量関数Ｐｒ（ε）は、特別な問題に関する目下の事前知識のすべてをコード化する。この特別な場合、これは、マージング処理特性についての知識若しくは確信を表わす（ホームビデオクラスタは、たいてい２，３個のショットのみからなる）。探究可能な多種多様な解決法が存在する。

－最も単純な想定は、Ｐｒ（ε＝０）＝Ｐｒ（ε＝１）＝１／２であり、ＭＡＰ判定基準をＭＬ判定基準に変える。

－事前確率自体は、訓練用データからＭＬ推定できる（Duda他によるPattern Classification, op.cit.参照）。Ｎが独立であると想定し、事前確率のＭＬ推定量が、

【００６１】

【数２２】

$$\Pr(\varepsilon = e) = \frac{1}{N} \sum_{k=1}^{N} \iota(e,k)$$

であることを示すことは明瞭である。ここで、

【００６２】

【外２】

$$\iota(e,k)$$

は、ｋ番目の訓練用サンプルが、ε＝ｅ，ｅ∈｛０，１｝により表現されるクラスに属する場合には１であり、それ以外の場合には０である。他言すると、事前確率は、利用可能な根拠（訓練データ）により決定される単なる重みである。

－マージングアルゴリズムに伴うダイナミックス（次の段落で示される）もまた、連続的な態様で事前知識に影響を及ぼす（より多くのセグメントが処理の終了時よりも開始時にマージされるだろうことが予測されている）。他言すると、事前確率は、この基本原理に基づいて動的に更新できる。

【００６３】５．ビデオセグメントクラスタリング

マージングアルゴリズムは、ビデオセグメントマージング段階１４で実行される。如何なるマージングアルゴリズムも３つの要素を必要とする。即ち、特徴モデル、マージングオーダー、及びマージング判定基準である（L. Garrido, P.Salembier, D.Garcia, “Extensive Operators in Partition Lattices for ImageSequence Analysis”, Sign. Proc., 66(2): 157-180,1998）。マージングオーダーは、どのクラスタが処理の各ステップで可能なマージングのための探査されるべきかを決定する。マージング判定基準は、マージングの可否の判断を行う。各クラスタの特徴モデルは、マージングが発生した場合には更新されるべきである。本ビデオセグメントクラスタリング法は、前の段落での改良型の統計的セグメント間モデルに基づいて、この一般的な定式化を利用する。本アルゴリズムにおいて、クラス条件が、マージング判定基準及びマージングオーダーの双方を定めるために用いられる。

【００６４】マージングアルゴリズムは、近接度グラフ及び階層的キューを用いて効率的に実行でき、優先順位付けした処理を可能とする。処理されるべき要素は、優先度を割り当てられ、それに従ってキューに案内される。このとき、各ステップで抽出された要素は、最も高い優先度を持つ要素である。階層的キューは、今日では、数学的形態学における従来的なツールである。その使用は、C.Chou及びC.Brownによる“The Theory and Practice of Bayesian Image Labeling”,IJCV,4,pp.185-210,1990にて、Highest Confidence First（ＨＣＦ）最適化法による、ベイズの画像解析で初めて開示される。その概念は、直感的なアピーリングである。即ち、決定

は、最も高い確実性を持つ情報片に基づいてなされるべきである。近年では、同様な定式化が、形態学処理において見受けられる。

【００６５】図２に示すように、セグメントマージング方法は、２つの段階、即ちキュー初期化段階２０とキュー更新／枯渇段階３０とを含む。マージングアルゴリズムは、バイナリベイズクラス分類器を構成し、このとき、マージングオーダーは、Highest Confidence First（ＨＣＦ）の変形により決定され、Maximum a Posteriori（ＭＡＰ）判定基準がマージング判定基準を定める。

【００６６】＜キューの初期化＞本処理の開始時（２２）、ショット間特徴x_{ij}は、ビデオ内のすべての隣接するショットの対に対して計算される。各特徴は、対応する対のショットをマージングする確率に等しい優先度Ｐｒ（$\varepsilon=1 \mid x_{ij}$）によりキューに案内される（２４）。

【００６７】＜キューの枯渇／更新＞優先度の定義は、最も高い確実性のある対のセグメントで常に決定をなすことを可能とする。キューが空になるまで（３２）、手順は次のとおりである。

１．要素抽出段階３４では、キューから要素（対のセグメント）を抽出する。この要素は、最も高い優先度を持つ。

２．ＭＡＰ判定基準を対のセグメントをマージするために適用する（３６）。即ち、

【００６８】

【数２３】

$$p(x_{ij} \mid \varepsilon=1)\Pr(\varepsilon=1) > p(x_{ij} \mid \varepsilon=0)\Pr(\varepsilon=0)$$

３．セグメントがマージされた場合（経路３８は、仮説H_1の適用を示す）、セグメントモデル更新段階４０で、マージされたセグメントのモデルを更新し、次いで、キュー更新段階４２で新たなモデルに基づきキューを更新し、ステップ１に行く。その他の場合、セグメントがマージされない場合（経路４４は、仮説H_0の適用を示す）、ステップ１に行く。

【００６９】対のセグメントがマージされたとき、新たなセグメントs_i'のモデルが

【００７０】

【数２４】

$$m_{i'} = (card(s_i)m_i + card(s_j)m_j)/(card(s_i)+card(s_j))$$

$$b_{i'} = \min(b_i, b_j)$$

$$e_{i'} = \max(e_i, e_j)$$

$$card(s_{i'}) = card(s_i)+card(s_j)$$

によって、更新される。

【００７１】（新たな）マージされたセグメントのモデルを更新した後、４つの機能が、キューを更新するために実現される必要がある。即ち、

１．元々の個々の（現段階では、マージされている）セグメントを伴うすべての要素のキューからの抽出。

２．更新モデルを用いた新たなセグメント間特徴ｘ＝（α、β、τ）の比較。

３．新たな優先度（$\varepsilon=1 \mid x_{ij}$）の計算。

４．新たな優先度に従った要素のキューへの挿入。

多くの上述の方法とは異なり、この定式化は、如何なる経験的なパラメータ決定を必要としないことに注目すべきである。

【００７２】マージングシーケンス、即ち、対のビデオセグメントの連続的なマージングを備えたリストは、階層を生成するために記憶及び利用される。更に、可視化及び操作に対して、マージングアルゴリズムで階層的キューを採用した後、更なるビデオセグメントのマージングにより、単一のセグメント（全体のビデオクリップ）に収束する完全なマージングシーケンスが確立される。マージングシーケンスは、次いで、可視的なコンテンツの階層的表現のための効果的な構造であると知られており、ユーザ対話に対する開始点を付与する、分割ツリー１８（図１）により表現される。

【００７３】６．ビデオ階層可視化

ツリー表現段階５０の一例が図５に示される。解析されたホームビデオのツリー表現を表示するためのインターフェースのひな形（プロトタイプ）は、各セグメントから抽出されたフレームである、キーフレームに基づいてよい。自動的に生成されたビデオクラスタの操作（修正、像拡大、再組織化）を、クラスタ再生や他のＶＣＲ能力と共に可能とする機能のセットは、上記表現に適用されてよい。ユーザは、このツリー表現を用いてビデオを解析し、プレビュークリップを取り出し、ビデオ編集をしてもよい。

【００７４】実際に評価された特性を用いたキューをベースとした方法は、バイナリ検索ツリーを用いて非常に効率的に実現でき、挿入、消去及び最小／最大位置特定の操作が明快である。本発明の好ましい実施例では、その実現は、L.Garrido, P.Salembier及びL.Garciaによる“Extensive Operators in Partition Lattices forImage Sequence Analysis”, Signal Processing, (66), 2,1998,pp.157-180の記述に関連する。

【００７５】マージングシーケンス、即ち、対のビデオセグメントの連続的なマージングを備えたリストは、階層を生成するために記憶及び利用される。階層の第１レベル５２は、ビデオセグメント化段階１０により付与される個々のセグメントからのキーフレームにより定義される。階層の第２レベル段階５４は、セグメントマージング段階１４で使用されたアルゴリズムにより生成されたクラスタからのキーフレームにより定義される。

【００７６】可視化及び操作に対して、マージングアルゴリズムで階層的キューを採用した後、更なるビデオセグメントのマージングにより、単一のセグメント（即ち、全体のビデオクリップを表現するキーフレーム段階

５６）に収束する完全なマージングシーケンスが確立される。全体のビデオクリップは、それ故に、階層の第３レベルを構成する。マージングシーケンスは、次いで、可視的なコンテンツの階層的な表現のための効果的な構造であると知られている、バイナリ分割ツリー（ＢＰＴ）により表現される。ＢＰＴでは、各ノード（リーフを除いて、初期のショットに対応する）は、２つの子を有する。（P.Salembier, L.Grrido, “Binary Partition Tree as an Efficient Representation for Filtering, Segmentation, and Information Retrieval”, IEEE Intl. Conference on Image Processing, ICIP ‘98, Chicago, Illinois, October 4-7,1998）また、ＢＰＴは、ユーザ対話のためのツールを確立するための開始点を提供する。

【００７７】ツリー表現は、自動的に生成されたビデオクラスタの可視化及び操作（検証、訂正、像拡大、再組織化）のための使いやすいインターフェースを提供する。ホームビデオコンテンツの一般性及び多様なユーザの好みの下で、人手によるフィードバック機構は、ビデオクラスタの生成を改善し、更にユーザのビデオに何かを実際に施しうる可能性をユーザに与えるだろう。

【００７８】マージング処理のツリー表現５０を表示するための単純なインターフェースでは、実行プログラムは、マージングシーケンスを読み出し、各セグメントから抽出されたフレームによりシーケンスの各ノードを表現する、バイナリツリーを確立するだろう。ランダムフレームは、ツリーの各リーフ（ショット）を表現する。各親ノードは、より少ないショット数の子のランダムフレームにより表現される。（用語“ランダム”は、“キーフレーム”を選ぶ際に一切の作業を要しないので、“キーフレーム”に代わって好ましいことに注意されたい。）注記するに、図５に示す表現は、また、マージング処理を可視化し、多数のクラスタを特定するのに有用であり、若しくは、ショット数が少ないときの一般的な表示のために有用であるが、それは、元のショット数が大きいときに非常に深くなることができる。

【００７９】インターフェースの第２バージョンは、階層の３つのレベルのみ；即ち、ツリーのリーフ、確率的マージングアルゴリズムの結果として得られるクラスタ、及び完全ビデオのノードのみを表示しうる。操作のモードは、マージングシーケンスの対話的な再組織化を可能とすべきであり、ユーザが自由にクラスタ間でのビデオセグメントの交換や、複数のビデオクリップからのクラスタの結合等をできるようにする。ツリーノードをクリックしたときのプレビューシーケンスの再生及びＶＣＲ能力のような、他の所望の機能を備えたインターフェースの統合は、当業者にとって明らかである。

【００８０】本発明は、好ましい実施例を参照して言及されてきた。しかしながら、変形及び修正が当業者により本発明の観点から逸脱することなくなされうることを理解すべきである。本発明の好ましい実施例は、消費者ホームビデオを例に言及されているが、本発明は、デジタル映画の要約及びストーリーボード化、ニュース及び製品関連のインタビューからのビデオ資材の組織化、動きが伴うヘルスイメージングアプリケーション等を含むがこれに限定されない、他のアプリケーションに容易に適用可能であることを理解すべきである。

【図面の簡単な説明】

【図１】本発明によるビデオ構造化の基本的な概要を示すブロック図である。

【図２】図１に示すビデオセグメントマージング段階のフローチャートである。

【図３】消費者画像の集合に対する消費者ビデオショット持続時間の分散プロット図である。

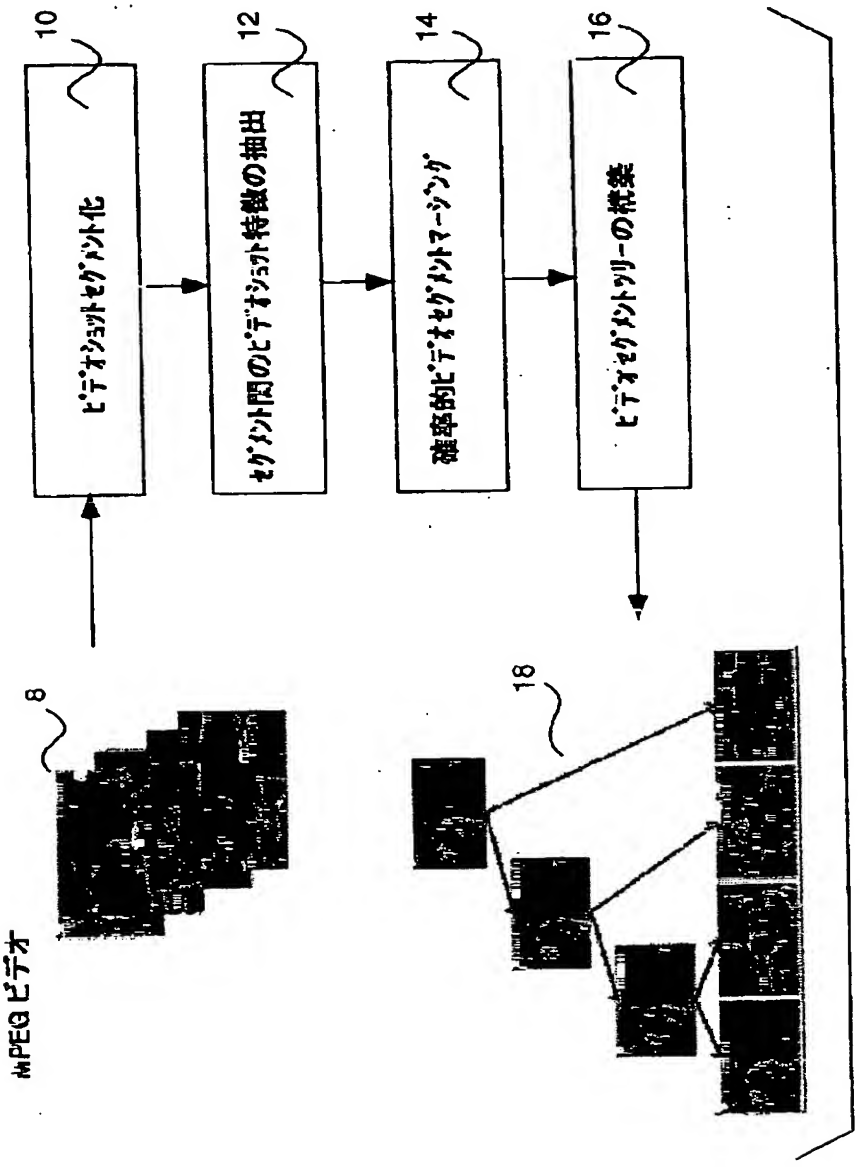
【図４】ホームビデオから抽出されたラベル付きのセグメント間特徴ベクトルの点在プロット図である。

【図５】典型的なホームビデオからのキーフレームのツリー表現を示す図である。

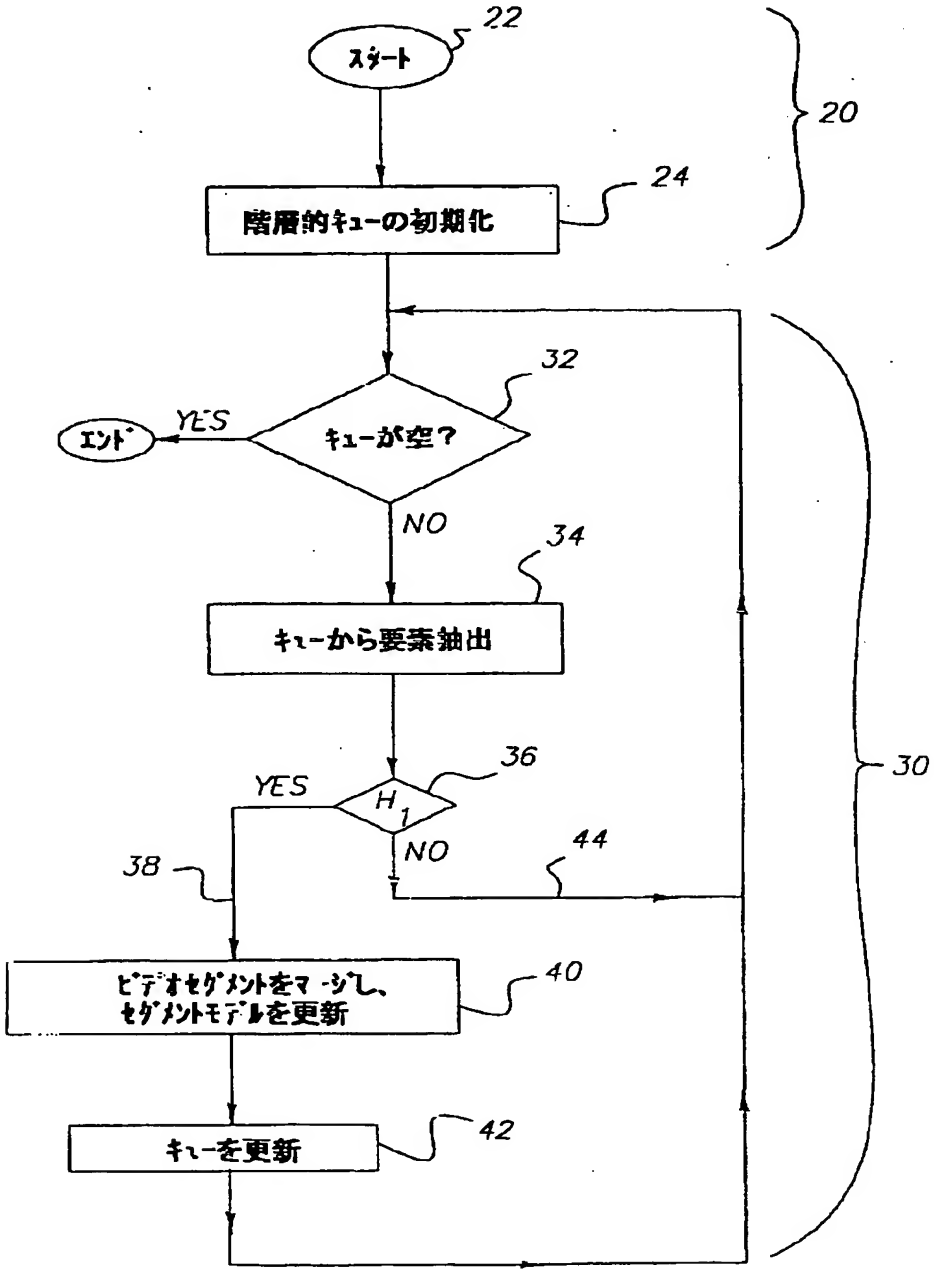
【符号の説明】

８　ビデオフレーム
１０　ビデオセグメント段階
１２　ビデオショット特徴抽出段階
１４　ビデオセグメントマージング段階
１６　ビデオセグメントツリー構築
１８　バイナリ分割ツリー
２０　キュー初期化段階
２２　処理の開始
３０　キュー枯渇／更新段階
３２　キューの空を判断
３４　要素抽出段階
３６　ＭＡＰ判定基準適用
３８　H_1 に対する経路
４０　セグメントモデル更新段階
４２　キュー更新段階
４４　H_2 に対する経路
５０　ツリー表現
５２　第１レベル
５４　第２レベル
５６　全体のビデオクリップ

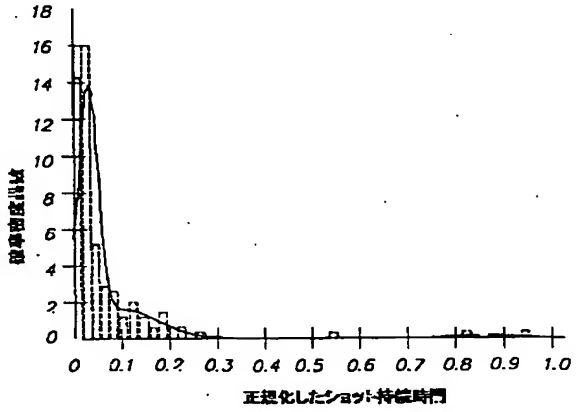
【図1】


10
ビデオショットセグメント化
12
セグメント間のビデオショット特徴の抽出
14
確率的ビデオセグメントマージング
16
ビデオセグメントツリーの構築
8
MPEG ビデオ
18

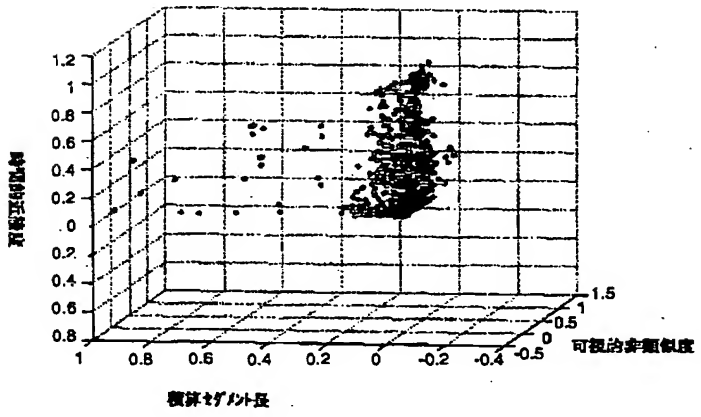
【図2】


22
スタート
20
24
階層的キューの初期化
32
YES
エンド
キューが空？
NO
34
キューから要素抽出
36
YES
H_1
NO
44
38
30
ビデオセグメントをマージし、
セグメントモデルを更新
40
キューを更新
42

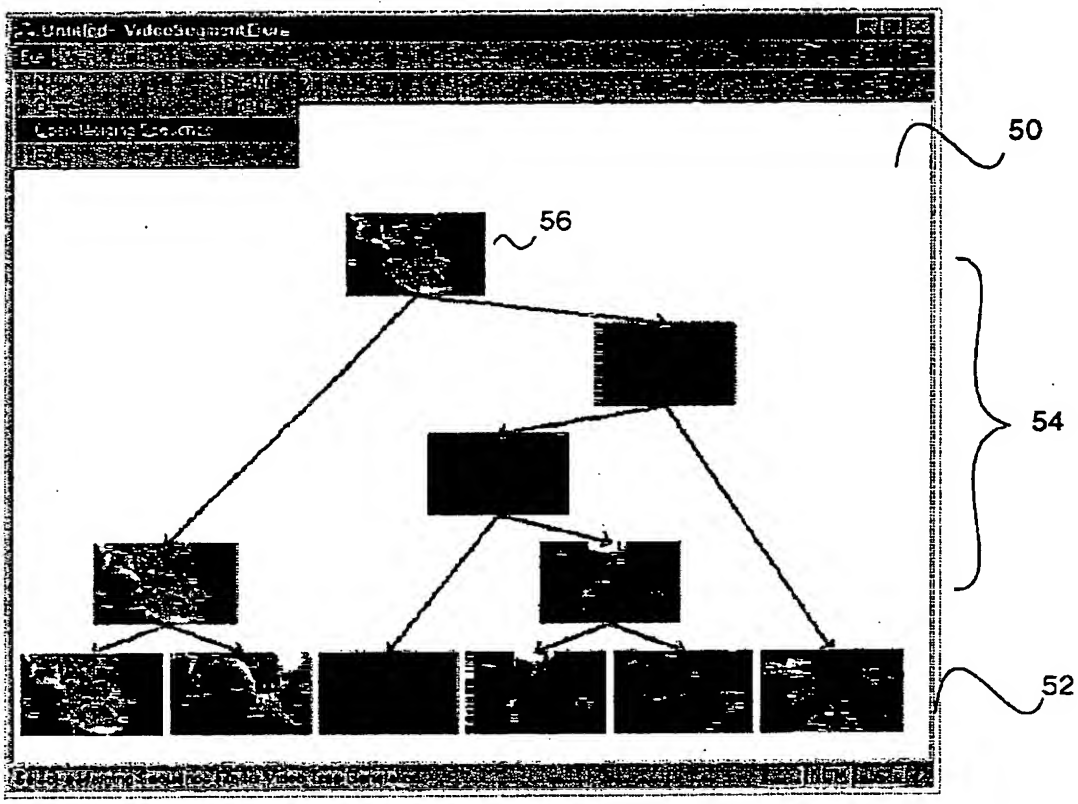
【図3】


18
16
14
12
10
8
6
4
2
0
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

【図4】


1.2
1
0.8
0.6
0.4
0.2
0
0.2
0.4
0.6
0.8
1 0.8 0.6 0.4 0.2 0 -0.2 -0.4
1.5
1
0.5
0
-0.5

【図5】



フロントページの続き

(72)発明者 ダニエル ガティカーペレス
アメリカ合衆国 ワシントン 98115 シアトル サーティース・アヴェニュー・ノースイースト 6053

Fターム(参考) 5C052 AB02 AC08 DD04 EE03